

Nociones sobre grafos

Definiciones

Grafo: par ordenado (V, E) donde

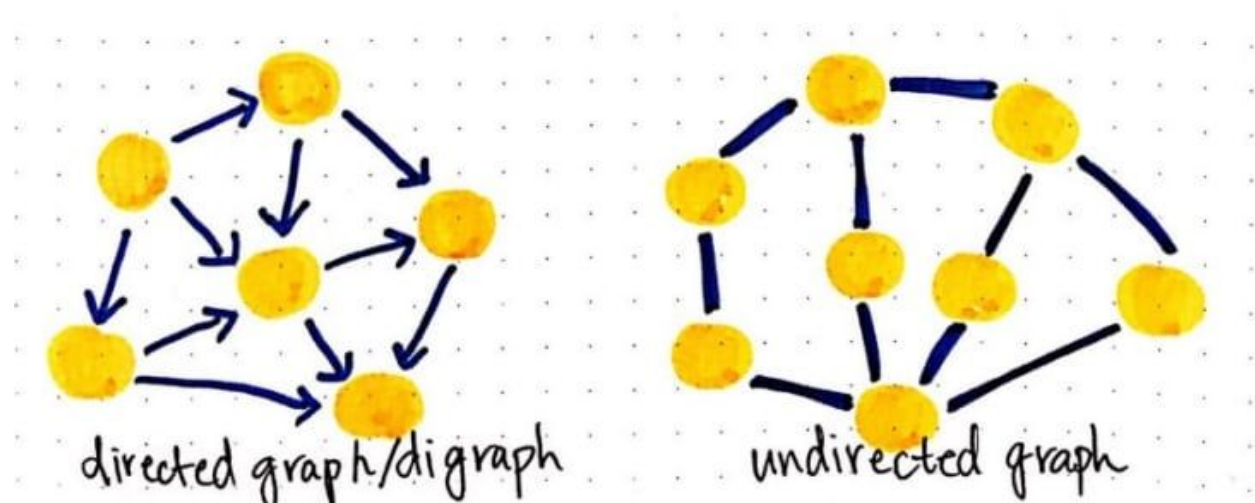
- V es el conjunto de vértices y
- E es el conjunto de aristas, donde una arista es
 - un conjunto de 2 elementos de V (grafo no dirigido)
 - un par ordenado de 2 elementos de V (grafo dirigido)

Si hay una arista entre v_1 y v_2 , decimos que los nodos son **vecinos** o **adyacentes**.

Ejemplos

GND: Facebook - **a** es amigo de **b** y **b** es amigo de **a**

GD: Twitter - **a** sigue a **b**



<https://dev.to/vaidehijoshi/a-gentle-introduction-to-graph-theory>

Grafo conexo: siempre hay un camino cualquier par de vértices.

Componente conexas: un subgrafo en el cual siempre hay un camino entre cualquier par de vértices.

Grafo completo: grafo en el cual todo par de vértices está conectado por una arista.

Clique: subgrafo completo

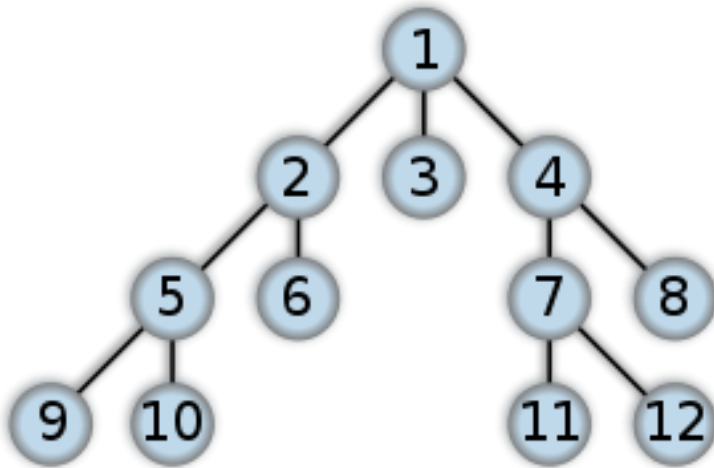
Grafo ponderado (o con peso): cada arista tiene un valor asociado.

Árbol es un tipo especial de grafos **conexo** y **sin ciclos**. Un **bosque** en el cual todas sus componentes conexas son árboles.

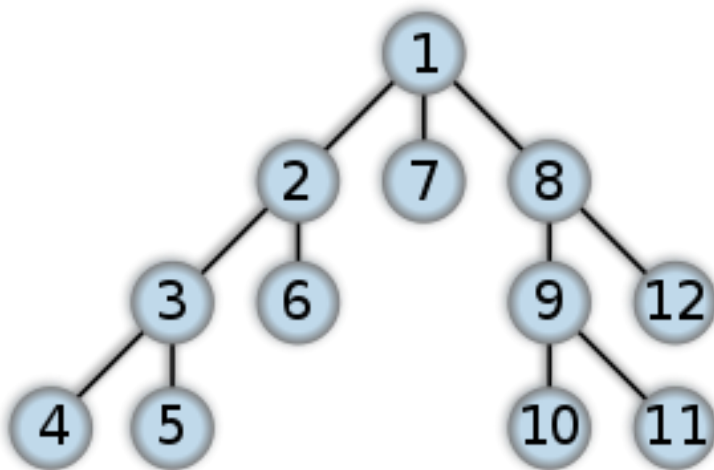
https://es.wikipedia.org/wiki/Anexo:Glosario_de_teor%C3%ADa_de_grafos

Algoritmos

Breadth-first search (BFS): Dado un vértice, primero se visitan todos los vecinos y luego los vecinos del siguiente nivel (cola).

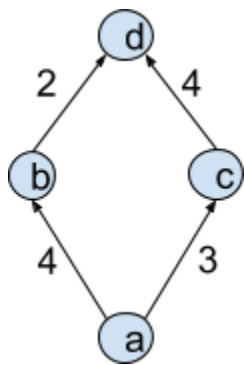


Depth-first search (DFS): Dado un vértice, se explora hasta la máxima profundidad antes de volver atrás (pila).

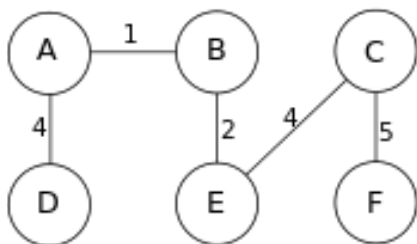
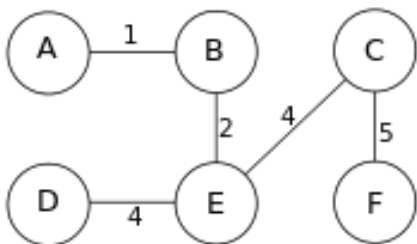
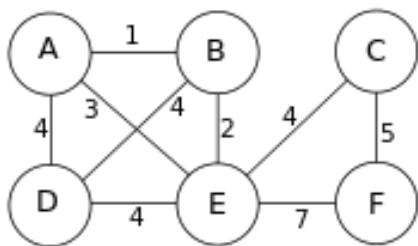


Shortest path first (SPF): Algoritmo de Dijkstra para encontrar el camino más corto entre 2 nodos.

Nota: El camino más corto no se obtiene de elegir la arista con menor peso en cada paso. En el siguiente grafo, ver el camino más corto en *a* y *d*.



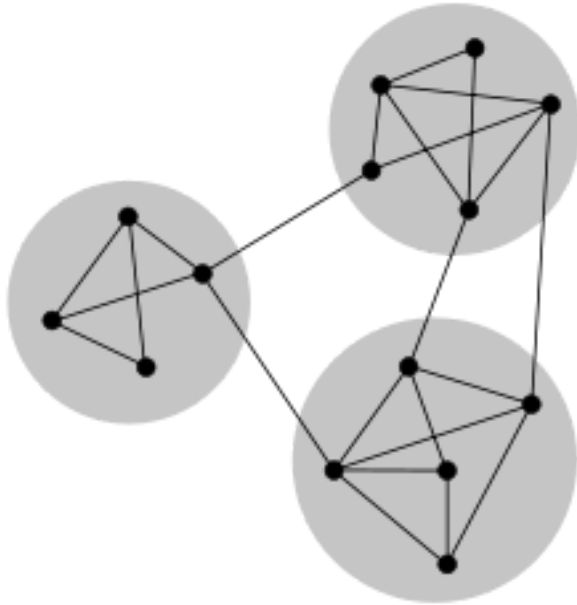
Minimum Spanning Tree (MST): En un grafo conexo, con pesos y no dirigido, un MST es un subconjunto de las aristas sin ciclos que conecta todos los vértices con el mínimo peso posible. Nótese que no es único.



Los dos árboles debajo son ejemplos de la no unicidad de un MST. Ejemplos de algoritmos de MST son el de Prim y Kruskal.

Comunidades

Comunidades: un par de nodos tiene mayor probabilidad de estar conectado si ambos son miembros de la(s) misma(s) comunidad(es), y menor probabilidad de estar conectado si no comparten comunidades.



Redes de Mundo Pequeño: la mayoría de los vértices no son vecinos, pero es probable que

1. los vecinos de un vértice sean vecinos entre sí
2. la mayoría de los nodos se pueden alcanzar desde otro en pocos pasos.

Grado: número de arcos con los que se conecta el nodo.

Cercanía: el promedio de distancia por el camino más corto al resto de nodos del grafo

Vector propio: medida de la influencia de un nodo en un grafo. **PageRank** es una variante de esto. Se puede usar una variante de esto donde la inicialización de los valores de los nodos no es aleatoria sino que se asigna a los nodos valores de importancia iniciales según el conocimiento de un experto del dominio. Si se usan estas medidas con flujo (representado por la direccionalidad de relaciones asimétricas como retweet, mención, etc.) se pueden detectar nodos **influyentes**.

Mediación (*betweenness*): número de veces que un nodo se encuentra en el camino más corto entre otros dos nodos

https://es.wikipedia.org/wiki/Estructura_de_comunidades

https://docs.google.com/document/d/1sibJmk_zYjMOIfGvdgQ7O23ccSTUpWV6lvGnZhDDBIg/e/dit#

Segmentación: Agrupar nodos usando atributos propios de los nodos.

Clustering: Agrupar nodos usando atributos de las relaciones.

